Miara informacji a jej znaczenie

Jacek Wolski

1 Wstęp

Od pulsarów i odległych gwiazd, poprzez geny żywej istoty lub skamieliny, do znaku drogowego i okrzyku ostrzegawczego, wszechświat jest wypełniony informacją. Niektórzy sądzą, że obok materii i energii jest to trzecia forma przejawiania się bytu. Rozwój nauk informacyjnych potwierdza to przekonanie. Pojęcie informacji od kilkudziesięciu lat jest poddawane rozmaitym badaniom. Matematycy, logicy, ale także ekonomiści i inżynierowie proponują różne teorie informacji. Fenomen ten inspiruje do badań między innymi ze względu na to, że informacja może wpłynąć na bieg historii lub zmienić przekonania narodów.

Na dzień dzisiejszy można wskazać kilkanaście różnych teorii informacji będących przedmiotem badań przedstawicieli filozofii lub różnych dziedzin nauki. Spośród wszystkich koncepcji warto zatrzymać się nad kilkoma z nich, ponieważ opisują główne linie rozwoju badań nad pojęciem informacji. Zostaną wyróżnione następujące teorie:

- matematyczna teoria komunikacji, jako przykład analizy informacji w kategoriach komunikacji;
- algorytmiczna teoria informacji opisująca złożoność informacji;
- teoria informacji odwołująca się do logiki indukcyjnej;
- teoria informacji odwołująca się do logiki epistemicznej;
- standardowa koncepcja informacji;

Zwrócimy uwagę na dwa wykluczające się typy opisów zjawiska informacji:

- szacunkowy, oparty na ilościowej ocenie, i
- semantyczny, opisujący znaczenie informacji.

2 Matematyczna Teoria Komunikacji

Literatura opisująca zagadnienie informacji wskazuje na Claude E. Shannona jako twórcę pierwszej teorii informacji. W latach 40-tych zbadał on bowiem zagadnienie komunikacji, czego wynikiem jest opublikowany w 1948 roku artykuł "Mathematical Theory of

Communication".[1] W artykule tym Shannon przedstawia matematyczny model przesyłania informacji przy pomocy środków komunikacji takich jak telegraf czy nadajnik radiowy. Praktyczne problemy, z jakimi stykał się autor w swojej pracy inżyniera, wymusiły ograniczenie zagadnienia do ściśle opisanego modelu. Model ten jest matematyczną idealizacją komunikacji. Zamiast analizować każdy możliwy sposób komunikacji, przykładowo akty mowy, sygnalizację świetlną czy łączność telegraficzną, autor decyduje się opisać pewien idealny system komunikacji. System taki składa się z następujących elementów:

- źródło informacji - element, który produkuje wiadomości, nadawca wiadomości,
- nadajnik, który przetwarza wiadomość przy pomocy pewnej funkcji kodującej na sygnał, który może być przesyłany poprzez kanał,
- kanał, którego zadaniem jest przenoszenie sygnału przy pomocy medium, na przykład przewody telefoniczne lub fale radiowe,
- odbiornik, który przetwarza sygnał na wiadomość za pomocą odwrotnej funkcji dekodującej, dokonując rekonstrukcji wiadomości,
- odbiorca, który jest odbiera wiadomość.[2]

Dla potrzeb omawianej modelowej sytuacji komunikacji, Shannon zakłada, że istnieje zbiór możliwych wiadomości, spośród których wybierana jest dokładnie jedna do przesłania. Wiadomość może być dowolnym (niekoniecznie sensownym) ciągiem znaków. Znaczenie wiadomości zostaje pominięte, ponieważ dla badanego problemu przesyłania informacji jest ono nieważne. Model komunikacji rozpatrywany jest przez Shannona w dwóch wersjach. W pierwszej z nich, pominięty zostaje problem szumów (zakłóceń), które są generowane poza modelem komunikacji. Jest to sytuacja idealna, nie występująca w rzeczywistości. W drugiej z nich, autor wprowadza pojęcie szumu i bada jego wpływ na możliwości przesyłania wiadomości, w tym na jakość wiadomości wobec poziomu szumu.

Kanał komunikacji może być opisany przez pojęcie pojemności. Pojemność jest parametrem określającym maksymalną miarą szybkości nadawania sygnału przez dany kanał. Przyjmując dla każdego symbolu jego określony czas zajmowania kanału komunikacyjnego, Shannon proponuje wzór określający pojemność kanału:

1 (Shannon 1948)
2 (Shannon 1948, s. 381)

(C)
$$C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T}$$

gdzie N(T) oznacza liczbę dozwolonych sygnałów w czasie T.[3] Jednostką pojemności jest bit na sekundę.4

W kolejnych częściach swego artykułu Shannon odpowiada na następujące pytania:

- czym jest źródło informacji,
- jakiej jakości jest informacja produkowana przez badane źródło,
- jaki jest wpływ czynników zewnętrznych na przesyłane komunikaty.

Ze względu na opisywane zagadnienie informacji, najbardziej interesującą częścią wspomnianej pracy jest ta, która dotyczy niepewności i entropii, bowiem związana jest bezpośrednio z opisem informacji. Shannon zadaje sobie bowiem pytanie, w jaki sposób zmierzyć niepewność jednego ze zdarzeń zawartych w zbiorze możliwych zdarzeń, gdzie dla każdego zdarzenia znane jest prawdopodobieństwo jego wystąpienia. Funkcja opisująca odpowiedź, która winna dostarczać takiej wiedzy, winna być wedle autora ciągła i monotoniczna, jeśli wszystkie zdarzenia są tak samo możliwe. Funkcja taka winna także spełniać warunek zależności wartości od sumy zdarzeń o ile wybór może być podzielony na dwa podobne wybory.[5] Opierając się na wcześniejszych badaniach w zakresie termodynamiki, autor proponuje do tej roli funkcję entropii *H*, która spełnia wymienione warunki i definiują ją za pomocą wzoru:

(H)
$$H = -K \sum_{i=1}^{n} p_i \log_2(p_i)$$

gdzie *K* jest dodatnią stałą. Jednostką tak definiowanej miary entropii jest bit na symbol. Entropia *H* jest proporcjonalna do ilości informacji produkowanej przez źródło. Opisuje średni poziom niesionej informacji przez ciąg symboli. Wzór (H) jest konsekwencją uogólnienia wyników analizy statystycznego prawdopodobieństwa wystąpienia symbolu.

---

3 (Shannon 1948, s. 382)
4 Jednostka miary zależy od podstawy logarytmu. Jeśli podstawa logarytmu wynosi 2, to wówczas jednostką jest bit na symbol (odpowiednio 10 - dit na symbol, e – nit na symbol). 8 bitów jest nazywane bajtem. Rozróżnienie jednostek obowiązuje także w dalszej części artykułu.
5 (Shannon 1948, s. 389)

Bliższa analiza wprowadzonego pojęcia entropii wskazuje na poprawność powyższego ujęcia. Przykładowo miara entropii jest równa zeru, o ile wszystkie zdarzenia, prócz jednego, są niemożliwe. Podobnie intuicyjnie entropia jest największa w systemach, gdzie prawdopodobieństwo wyboru jednego z symboli jest stałe i równe dla każdego z symboli.

Z punktu widzenia praktycznego zastosowania matematycznej teorii komunikacji, najważniejsze są dwa twierdzenia opisujące zależność pomiędzy entropią wiadomości a pojemnością kanału. Shannon dowodzi, że dla źródła o określonej entropii *H*, kanale o pewnej pojemności *C* i zerowym poziomie szumu, wiadomość może być przesłana przy średniej prędkości nie wyższej niż iloraz pojemności i entropii.[6] Podobne twierdzenie opisuje zależność pomiędzy pojemnością a entropią, przy założeniu danego poziomu szumu. Otóż jeśli entropia *H* jest niewiększa niż pojemność kanału, to istnieje system kodowania taki, że wiadomość może być transmitowana przez kanał z małą częstością błędów. W przeciwnym wypadku, jeśli entropia *H* jest większa niż pojemność kanału *C*, to jest możliwe takie kodowanie wiadomości, że ewentualne błędy będą mniejsze niż $H\text{-}C+\varepsilon$, gdzie ε jest pewną wielkością stałą.[7]

Jako egzemplifikację matematycznej teorii komunikacji (w dalszym ciągu będę się posługiwał skrótem MTC) rozważmy skończony zbiór symboli

$$S=\{A_{1,}\ A_{2,}\ A_{3,}\ A_4\}$$

Jeśli dla każdego symbolu zostanie założony ten sam poziom prawdopodobieństwa w wysokości 0,25, to wówczas entropia obliczona na podstawie wzoru (H) wyniesie 2 bity na symbol. Natomiast jeśli dla każdego symbolu zostanie przyjęty inny poziom prawdopodobieństwa, przykładowo:

p(A1)=0,6

p(A2)=0,1

p(A3)=0,2

p(A4)=0,1

entropia wyniesie 1,57 bita na symbol.

6 (Shannon 1948, s. 395)
7 (Shannon 1948, s. 401)

Jeśli dla każdej jednostki czasu *T* równej 1 sekundzie, liczba dozwolonych sygnałów w jednostce czasu jest identyczna i wynosi 2 symbole, wówczas pojemność kanału *C* wynosi, zgodnie ze wzorem (C), 1 bit na sekundę. Zatem średni poziom prędkości przesłania sygnału, który jest losowo wybraną wiadomością ze zbioru symboli, wyniesie 2 symbole na sekundę. Natomiast w drugim rozpatrywanym przypadku, gdy wiadomość jest układana z symboli występujących z pewną częstością, średnia prędkość przesyłania wiadomości wyniesie 1,57 symboli na sekundę.

Podsumowując, Shannon proponuje matematyczny opis komunikacji. Informację postrzega w kategoriach własności symboli użytych do przesłania wiadomości. Znaczenie informacji jest umyślnie pominięte, zaś ilość informacji jest odwrotnie proporcjonalna do prawdopodobieństwa wystąpienia danego symbolu, co w konsekwencji sprowadza się do tego, że im mniejsze jest prawdopodobieństwo przesyłanej wiadomości, tym większą informację niesie taka wiadomość. W omawianej koncepcji Shannona nie można stwierdzić explicite zależności pomiędzy ilością informacji a jej znaczeniem. Jeśli zostanie wykazany stosunek pomiędzy prawdopodobieństwem informacji a jej znaczeniem, wówczas pośrednio można oceniać relację ilości i znaczenia informacji.

3 Algorytmiczna teoria informacji

MTC nie jest jedyną teorią, która szacuje liczbowo wartość informacji. Opierając się na wynikach osiąganych w informatyce, niektórzy zaproponowali inne podejście do problemu wartości informacji. Ich badania są podstawą algorytmicznej teorii informacji. Kluczową ideą tej koncepcji jest założenie, że teoria, która wyjaśnia zjawisko *X*, jest swego rodzaju programem komputerowym, który oblicza *X*. Zatem taki program musi być w pewnym sensie mniejszy niż samo zjawisko, które wyjaśnia. Koncepcja informacji opiera się na takim stwierdzeniu i wynikach osiągniętych przy badaniu złożoności w informatyce.

Algorytmiczna teoria informacji definiuje złożoność informacji lub inaczej zawartość informacji *X* w kategoriach wielkości H(X) czyli najmniejszego możliwego programu komputerowego, który oblicza *X*. Analogicznie, H(X) może być traktowane jako złożoność najprostszej teorii wyjaśniającej *X*.[8]

---

8 por. (Chaitin 2003)

Złożoność obliczeniowa może być opisana w każdym znanym języku programowania, przykładowo *C*, *LOGO*, *LISP* czy *PASCAL*, o ile można w nim zbudować uniwersalną maszynę Turinga. Zależność ta wynika z tego, że złożoność obliczeniowa algorytmu jest zwykle opisywana formalnie przy pomocy maszyny Turinga. Maszyna Turinga służy do formalnej symulacji pracy dowolnego komputera.

Najprostsza definicja wartości algorytmicznej informacji, tak zwanej złożoności Kołmogarowa, wyraża się wzorem:

(O) $$H(s) = \left(\min(l(y) + l(s)\right)$$

dla każdej funkcji opisanej jako s=p(y), gdzie *p* oznacza pewien program, który przetwarza dane wejściowe *y* na ciąg *s*, a następnie zatrzymuje się. Zgodnie z opisanym wzorem, złożoność obliczeniowa jest najmniejszą sumą długości programów l(s) i l(y).

Najprostszy program, który może obliczyć dany ciąg s, to jedna linia kodu zapisana na przykład w języku programowania *C*:

```
println(11111110);
```

Taki program działa bez danych wejściowych. Wypisuje żądany ciąg *s* i zatrzymuje się. Złożoność takiej informacji wynosi *H(s)=O(n)*. Zapis ten oznacza, że złożoność obliczeniowa jest proporcjonalna do długości informacji *s*, która jest przedmiotem obliczeń. W zależności od typu informacji można wskazać różne poziomy złożoności obliczeniowej.[9]

Warto zauważyć, że dany algorytm zapisany w postaci programu w jednym języku programowania może być przeniesiony, „przetłumaczony” na inny dowolny język programowania. Oznacza to, że każdy język może być wykorzystany do wyliczania złożoności, ponieważ złożoność takiego przekładu jest wielomianowa, a zatem wprost proporcjonalna do długości programu.[10]

Potoczne intuicje wiążą pojęcie informacji z pewnymi stanami zachodzącymi w świecie. Zwykle uznaje się, że informacja „mówi o czymś” czy „jest na jakiś temat”. Warto zauważyć, że informacja w omawianej koncepcji nie spełnia tego warunku. Wartość

---

9 por. (Grünwald, Vitanyi 2008)
10 (Markowsky 1996, s. 254)

informacji w tej koncepcji dotyczy tylko i wyłącznie informacji jako pewnego ciągu *s*, bez żadnych związków z treścią informacji.

Idea tej koncepcji, by przedstawiać wartość zawartości informacji, przy pominięciu znaczenia, jako wielkość najmniejszego programu, który oblicza daną informację, ma jedną wadę. Program, który drukuje dany ciąg i zatrzymuje się, może być różnej długości, przy czym nie ma żadnego algorytmu, który mógłby określić, że dany program jest najkrótszym z możliwych. Niektórzy, na przykład G. Chaitin, wprowadzają dodatkową wartość elegancji programu, który określa prostotę badanego programu. W ten sposób można prowadzić badania i szukać lepszych algorytmów i ich zapisu w formie programu, zgodnie z opisanym kryterium. Badanie i porównywanie algorytmów opiera się na ocenie ich prostoty, stopnia minimalizacji wykorzystania posiadanych zasobów lub szybkości działania w zależności od ilości danych wejściowych.[11]

Jako egzemplifikację złożoności informacji, należy rozważyć zbiór symboli

$$S = \{A_{1,}\ A_{2,}\ A_{3,}\ A_4\}$$

Informacją w tym przypadku może być ciąg

$$(A_4, A_1, A_2, A_3)$$

Program, który oblicza taką informacją może być zapisany w co najmniej dwóch postaciach w języku zbliżonym do *C*, gdzie zapis S(1) oznacza wybranie 1 elementu z czteroelementowej tablicy *S*. Zakładam, że tablica *S* została wcześniej zdefiniowana. W danym zbiorze *S* oznacza to element A_2.

Wersja 1

```
println(S(3),S(0),S(1),S(2));
```

Wersja 2

```
k=0;
for(i=0,3){
k=(k+3)mod 4;
println(S(k));}
```

11 (Markowsky 1996, s. 260)

W wersji 1 szukany ciąg jest wyświetlany na monitorze za pomocą funkcji *println*. Funkcja ta jest najprostszym sposobem na wysłanie ciągu znaków na ekran monitora. W wersji 2 algorytm jest oparty na pętli *for* i obliczeniach wskazujących kolejny wyświetlany elementu. Linia 1 określa parametry wejściowe. Linia 2-4 dotyczą pętli *for*. Linia 2 wskazuje jak długo będzie działał program tj. od 0 do 3 czyli pętla zostanie wykonana dokładnie 4 razy, a potem program się zatrzyma. Linia 3 oblicza wartość *k*, która w linii 4 jest argumentem instrukcji *S(k)* przywołującej wartość elementu *k* tablicy *S*.

Wielkość programu w wersji pierwszej wynosi 29 znaków. Przy założeniu, że jeden znak języka programowania wysokiego poziomu, takiego jak język *C*, jest kompilowany do dokładnie jednego bajtu programu, wielkość programu wynosi 29 bajtów. Wielkość programu w wersji 2 jest nieco dłuższa i wynosi 43 bajtów. W rezultacie oba programy, choć wypisują taką samą informację długości 4 bajtów, mają różną wielkość. Wersja 2, choć większa w stosunku do wersji 1, w prosty sposób może być dostosowana do wyświetlania dłuższej informacji. Złożoność obliczeniowa jest nadal rzędu O(n), ponieważ wielkość programu jest zależna od wielkości *n*, która opisuje długość badanego ciągu *s*.

4 Informacja w ujęciu Carnapa i Bar-Hillela

W latach 50-tych, R. Carnap, wraz z Y. Bar-Hillelem, prowadzili badania nad logiką indukcji. Indukcja, jako forma wnioskowania zawodnego, miała być sformalizowana na podobieństwo innych sposobów dowodzenia. Taki zakres badań był inicjowany rozwojem filozofii nauki.[12] Podstawą logiki indukcyjnej, zgodnie z wynikami badań Carnapa, miała być funkcja prawdopodobieństwa. W przeciwieństwie do koncepcji Shannona, który mówił o prawdopodobieństwie statystycznym, Carnap zaproponował prawdopodobieństwo polegające na stopniu potwierdzenia badanej hipotezy na podstawie posiadanych dowodów.[13] (Warto zauważyć, że R. Carnap w późniejszych latach zmienił swoje poglądy co do znaczenia prawdopodobieństwa)

Koncepcja informacji w logice indukcyjnej oparta jest na pewnym logicznym modelu świata. Carnap proponuje metodę obliczania prawdopodobieństwa potwierdzenia (konfirmacji) w prostej przestrzeni logicznej. Ogranicza on bowiem język indukcji do

12 (Carnap 1966, s. 248)
13 (Carnap 1963, s. 67)

klasycznych spójników takich jak negacja czy alternatywa i ściśle określanego zbioru nazw, który jednocześnie wyznacza zbiór przedmiotów w tym uniwersum oraz zbioru predykatów jednoargumentowych.[14]

Podstawowymi elementami świata są Q-predykaty czyli koniunkcje wszystkich jednoargumentowych predykatów, zwanych przez Carnapa prostymi. Różnią się one od siebie tylko liczbą zanegowanych predykatów. Każdy z predykatów prostych może być zdefiniowany jako alternatywa Q-predykatów. Ponieważ Q-predykaty obejmują wszystkie predykaty proste z danego zbioru, liczba wszystkich Q-predykatów wynosi:

$$k=2^{r},$$

gdzie *r* to liczba predykatów prostych.

Carnap wprowadza dodatkowo dwie konstrukcje, które opierają się na wprowadzonych Q-predykatach. W języku Carnapa można mówić o opisach świata (state descriptions) i opisach statystycznych (structure descriptions). Te dodatkowe konstrukcje są ważne dla obliczenia miary i zawartości informacji (content of information).[15]

Każdy opis świata jest koniunkcją *n* Q-predykatów, gdzie argumentem jest nazwa przedmiotu, *n* jest liczbą tych nazw. Przy czym należy zaznaczyć, że każdy człon koniunkcji opisuje inny przedmiot. Każdy opis charakteryzuje kompletnie świat. Zatem każde możliwe zdanie w języku *J* jest sprzeczne z opisem świata bądź wynika z tego opisu świata. Opis świata może być traktowany jako syntaktyczny korelat świata możliwego. Liczba opisów świata, opisanych światów możliwych, wynosi $k*n$, gdzie *k* jest liczbą Q-predykatów, a *n* jest liczbą nazw w badanym języku *J*. Opis statystyczny jest alternatywą podobnych opisów świata. Podobieństwo wynika z tego, że opisy świata mogą być pogrupowane wedle tego, jakie Q-predykaty są ujęte w danym opisie świata. Carnap nie wskazuje jednoznacznie jednego sposobu określania miary informacji. Wynika to stąd, że badacz ten najpierw określa istotne warunki, które winna spełniać miara informacji, a potem stara się je spełnić, zakładając Q-świat jako miejsce sprawdzania. Carnap uważa, że pojęcie miary informacji winno spełniać następujące warunki:

a) inf(j)=0, gdy *j* jest logicznie prawdziwe

b) inf(j)=1, gdy *j* jest logicznie fałszywe

c) $0 \leq \text{inf(j)} \leq 1$, jeśli *j* jest faktem

---

14 (Mortimer 1982, s. 64)
15 (Mortimer 1982, s. 66)

d) $\inf(i \vee j) \leq \inf(j) \leq \inf(i \wedge j)$

gdzie inf(j) oznacza miarę informacji zdania *j*.[16]

Spełnienie tych, a także innych warunków jest możliwe dla dwóch funkcji, które wzajemnie się uzupełniają i spełniają wszystkie warunki, które wynikają z pracy badawczej nad pojęciem informacji. Zależność pomiędzy tymi dwiema funkcjami jest opisana następująco:

(inf) $$\inf(i) = \log_2\left(\frac{1}{1 - cont(i)}\right)$$

gdzie *inf* oznacza funkcję miary informacji, natomiast *cont* opisuje funkcję miary zawartości informacji. Można opisać wartość funkcji cont(i) dla prostych zdań, każdego zdania złożonego z zdań prostych i przedstawione w dysjunkcyjnej lub koniunkcyjnej postaci normalnej i dla każdego Q-zdania. W tym ostatnim przypadku wartość ta wynosi:

(cont) $$cont(i) = 1 - \left(\frac{1}{2}\right)^r$$

gdzie *r* to liczba prostych predykatów.[17] Jak można zauważyć, miara zawartości informacji cont(i) określa średnią wartość każdego Q-zdania. W tej koncepcji język, ograniczony tylko do skończonego zbioru predykatów jednoargumentowych i skończonego zbioru nazw, powoduje, że nie można wskazać wyróżnionego zbioru zdań, które niosłyby więcej informacji. Oznacza to niezależność wartości informacji wobec potencjalnego znaczenia informacji.

Skoro inf(i) jest proporcjonalna do cont(i), to cont(i) spełnia te same warunki jak inf(i). Można wskazać tutaj na pewien paradoks polegający na tym, że wartość inf(i) wynosi 1, gdy *i* jest logicznie fałszywe, czyli gdy jest kontrtautologią, zatem zdania fałszywe zawierają więcej informacji niż zdania prawdziwe. Jest to sprzeczne z potocznymi intuicjami, które wyróżniają zdania prawdziwe jako ważniejsze, niosące więcej informacji. Jednak Carnap nie uważa takiej sytuacji za paradoksalną. Twierdzi

16 (Carnap, Bar-Hillel 1952, s. 11)
17 (Carnap, Bar-Hillel 1952, s. 17)

bowiem, że zdania fałszywe niosą zbyt dużo informacji, zbyt wiele mówią o świecie by być prawdziwe.[18]

Warto zauważyć, że prace Carnapa i Bar-Hillela obejmowały znacznie większy zakres zagadnień niż przedstawiany tutaj. Zaproponowali bowiem różne sposoby ujmowania wartości informacji, ze szczególnym naciskiem na miarę ilości informacji.

Celem Carnapa była formalizacja procesu indukcji, łącznie ze wskazaniem sposobu określania prawdopodobieństwa prawdziwości hipotezy na podstawie zebranych dowodów. Jednak przyjęcie ubogiego języka ogranicza potencjalne wyniki. Niewiele teorii naukowych operuje językiem, który składa się tylko ze skończonego zbioru predykatów jednoargumentowych i skończonego zbioru nazw wyznaczającego jednocześnie zbiór wszystkich przedmiotów uniwersum.

Rozważmy przykład. Niech będzie dany zbiór symboli *S*

$$S = \{A_{1,} A_{2,} A_{3,} A_4\}$$

Symbole te mogą oznaczać w omawianej koncepcji nazwy pewnego skończonego zbioru przedmiotów. Wówczas należy określić zbiór predykatów jednoargumentowych, który definiowałby język. Symbole te mogą oznaczać również cztery kolejne predykaty. Dla potrzeb przykładu symbole zbioru *S* oznaczają zbiór nazw. Istnieje tylko jeden predykat *P*. Zatem język *L* składa się dokładnie z 4 stanów atomowych, 2 Q-predykatów i 16 stanów świata. Wartość miary zawartości informacji wynosi, zgodnie z wzorem (cont):

cont(i)=0,5

Wartość miary informacji wynosi, na podstawie wzoru (inf):

inf(i)=1

W tej koncepcji nie zostały określone jednostki informacji. Wykorzystanie w definicji *inf* miary informacji logarytmu o podstawie 2 wskazuje, że można mówić w tym wypadku o bicie jako jednostce. W omawianym przykładowym języku, miara zawartości informacji rozkłada się równomiernie według Q-predykatów. Miara informacji jest proporcjonalna do zawartości. Carnap mówi tylko o ilościowym charakterze informacji. Pomija jej znaczenie.

18 (Carnap, Bar-Hillel 1952, s. 7-9)

5 Ogólna definicja informacji

Badania nad pojęciem informacji są prowadzone na różnych polach współczesnej logiki. Są oparte na wynikach rozwoju szczególnych gałęzi logiki, na przykład logiki sytuacyjnej.[19] Niektórzy badacze zaproponowali standardową definicję informacji (SDI – Standard Definition of Information) znaną również jako ogólna definicja informacji (GDI – General Definition of Information). Logika sytuacyjna winna dostarczać tej koncepcji zaplecza formalnego.[20] Definicja informacji w tej koncepcji brzmi następująco:

(D) *s* jest informacją, wtedy i tylko wtedy gdy:

- *s* zawiera *n* danych
- dane są dobrze uformowane
- dobrze uformowane dane mają znaczenie.

Dane, będące podstawą definicji (D), są prostymi elementami, częściami otaczającej podmiot rzeczywistości, które nie są jednostkowe. Dane te mogą być podstawą wrażeń zmysłowych. Mogą być także traktowane jako fenomeny świata. Ten typ danych związany jest z elementami świata, z którymi pewna osoba wchodzi w interakcje. Danymi są także różnice pomiędzy dwiema sytuacjami, które następują po sobie lub które współwystępują równocześnie. Brak jedności, który jest jedynym warunkiem bycia daną, może pojawiać się pomiędzy dwoma literami alfabetu. Oba scharakteryzowane typy danych dotyczą interpretacji pewnych doznań podmiotu w kategoriach danych.[21]

Pierwszy warunek definicji informacji (D) podkreśla tę cechę informacji, która polega na tym, że informacja bez danych nie jest informacją. Dane, o jakich mowa w definicji (D) są warunkiem wystarczającym i koniecznym informacji. Dane te mogą należeć do jednej z czterech kategorii:

1. dane pierwotne – powszechnie postrzegane podstawowe dane, które mogą być zawarte w różnych bazach danych takich jak tablice numerów czy w postaci zawartości książek w bibliotece,
2. metadane – określają ważne cechy danych pierwotnych potrzebne dla różnych systemów zarządzania bazami danych; należą do nich: miejsce przechowywania, dostęp, ograniczenia praw autorskich i inne,

19 por. (Barwise, Perry 1983)
20 (Floridi 2005a, s. 353)
21 por. (Floridi 2005b)

3. dane operacyjne – określane jako operacje całego systemu baz danych
4. dane derywacyjne – wszystkie te dane, które wynikają z powyższych trzech kategorii.[22]

Drugi warunek definicji (D) dotyczy składni danych. Informacja nie jest chaotycznie uporządkowanymi danymi. Dane są ułożone zgodnie z założoną dla danego kodu czy języka składnią. Pojęcie składni nie powinno być tutaj ujmowane lingwistycznie, lecz raczej w kategoriach formy lub zasad. Natomiast spełnienie trzeciego warunku definicji (D) powoduje, że dane składające się na informację muszą posiadać znaczenie semantyczne. Warunek ten można traktować w kategoriach denotacji i konotacji.

Tak opisaną definicję informacji można dodatkowo uszczegółowić wskazując na pewne dodatkowe cechy, które wskazują na neutralność wobec niektórych aspektów rzeczywistości. Można wyróżnić następujące typy neutralności danych:

- wobec typów,
- wobec klasyfikacji,
- wobec ontologii,
- neutralności genetyczna,
- neutralność aletyczna.

Pierwsza cecha danych, obojętność wobec typów, oznacza, że informacje mogą składać się z danych różnych typów czy rodzajów. Dane mogą być różne i różnie reprezentowane przez kod, sygnały, postrzeżenia czy związek pomiędzy nimi.

Druga cecha danych, neutralność taksonomiczna, stwierdza, że dane są bytami, które pozostają w relacji z innymi danymi. Dane nie występują samodzielnie, są powiązane ze sobą pewnymi związkami. Pusta strona na ekranie również jest daną, choć intuicyjnie wydaje się, że dopiero pierwszy znak, który się pojawi na skutek naciśnięcia klawisza klawiatury, niesie jakąkolwiek informacją. Bycie daną jest zewnętrzną cechą, a nie istotą czegokolwiek.[23]

Neutralność ontologiczna wskazuje na to, że informacja musi składać się z danych. W omawianej koncepcji odrzucona zostaje możliwość istnienia informacji bez danych.

22 (Floridi 2005a, s. 354)
23 (Floridi 2005a, s. 356)

Jeśli ujmować ten warunek materialistycznie, dane reprezentują fizyczny świat. Dane mogą być każdym elementem rzeczywistości.[24]

Neutralność genetyczna określa stosunek informacji wobec odbiorcy informacji. Dane niosą informacje niezależnie od istnienia odbiorcy czy nadawcy informacji. Zakodowany telegram, zawiera informację, choć jest ona jeszcze niezrozumiała.

Przedstawiona definicja informacji może zostać sformalizowana przy pomocy narzędzi logiki sytuacyjnej. Logika sytuacyjna zakłada, że fakty niosą informacje, a informacja jest opisana jako zespół danych. Logika sytuacyjna operuje modelem, którego częścią jest sytuacja jako pewien wydzielony fragment rzeczywistości. Logika ta jednak nie dostarcza opisu ilości informacji.[25]

Jako przykład standardowej definicji informacji, należy rozważyć zbiór symboli

$$S = \{A_{1,} A_{2,} A_{3,} A_4\}$$

gdzie każdy symbol oznacza pojedyncze datum, pojedynczy znak pochodzący z 4-znakowego alfabetu. Wówczas informacją można nazwać jedynie konstrukt, który składa się z co najmniej jednego symbolu, jednocześnie jest dobrze uformowany wedle określonej składni i ma znaczenie w szerokim sensie tego słowa. Aby przedstawić egzemplifikację tej koncepcji informacji, należy dysponować składnią i semantyką takiego alfabetu. Ta koncepcja koncentruje się na aspekcie znaczeniowym informacji i jej relacji ze światem. Zaproponowany model jednak nie uwzględnia ilościowego aspektu informacji. Można ewentualnie rozważać przeliczalny zbiór informacji wynikający ze zbioru danych. W koncepcji tej nie można wskazać zależności pomiędzy ilością a znaczeniem informacji.

6 Informacja w terminach logiki epistemicznej

Niektórzy badacze zaproponowali formalizacje modalności posiadania przekonań, klasycznie wyrażanej operatorem „wie, że".

Do opisu wiedzy w logice epistemicznej często korzysta się z modelu opracowanego w latach sześćdziesiątych przez S. Kripkego. Model ten wykorzystuje pojęcie światów możliwych do rozróżniania pomiędzy różnymi stanami epistemicznymi. Stany te obrazują wiedzę podmiotu.

24 (Floridi 2005a, s. 357)
25 por. (Israel, Perry 1990)

Model Kripkego M jest strukturą {S, V, R}, gdzie:

- *S* to niepusty zbiór światów możliwych
- *V* to funkcja interpretacji, która każdej zmiennej zdaniowej przypisuje podzbiór zbioru świata możliwego czyli zbiór tych światów możliwych, w których prawdziwe jest dane zdanie atomowe,
- *R* jest relacją dostępności (accesibility) zdefiniowaną na zbiorze *S*.

W takiej przestrzeni zostaje zdefiniowany operator $K\varphi$, który można wyrazić w postaci „wie, że φ". Precyzyjnie przy użyciu modelu Kripkego, definicja przyjmuje następującą postać:

$$M, s \models K\varphi \iff \forall t\, R(s,t) \rightarrow M, s \models \varphi$$

gdzie *s*,*t* są elementami zbioru *S*. Relacja *R* wyróżnia jeden świat możliwy, jako korelat semantyczny aktualnej wiedzy podmiotu.

Wiążąca jest w badanym modelu następująca definicja przykładowego spójnika negacji:

$$M, s \models \sim \varphi \iff \sim \big(M, s \models \varphi\big)$$

Minimalny system wiedzy K składa się ze wszystkich aksjomatów klasycznego rachunku zdań i jednego aksjomatu występującego we wszystkich tzw. systemach normalnych logiki, gdzie miejsce operatora konieczności zajmuje operator wiedzy *K*. Aksjomat ten brzmi następująco:

$$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$$

Jako reguły inferencji należy przyjąć reguły systemó*w* logiki modalnej z odpowiednią zmianą operatora konieczności na operator K.

Można dowieść, że taki system K jest niesprzeczny i zupełny. Dalsze uzupełnianie aksjomatów systemu K prowadzi do powstania systemów o strukturze odpowiednio T, S4 lub S5.[26]

Przedstawiony system *K* dotyczy tylko jednego agenta. Operator modalny *K* może być zdefiniowany w taki sposób, by określać stany epistemiczne dla skończonego zbioru

26 por. (Meyer 2001)

agentów. Wówczas należy omawiany system *K* uzupełnić w taki sposób, by jednoznacznie rozróżnić pomiędzy światami możliwymi opisującymi wiedzę danego agenta. Wówczas operator K_ix oznacza „ agent *i* wie, że *x*”.

Logika epistemiczna stanowi jednocześnie teorię informacji. Stany epistemiczne są stanami informacji, zatem w tej koncepcji wiedza i informacja są pojęciami równozakresowymi. W logice epistemicznej zostaje wprowadzone pojęcie informacji jako zasięgu, ponieważ każdy świat możliwy, opisujący wiedzę agenta, opisuje w swym zasięgu posiadane przez agenta informacje. Przekonania i informacje są wynikiem czynności, aktywności agenta. Komunikowanie się z innymi, obserwacja otaczającego świata czy rozumowania traktowane jako łączenie informacji pochodzących z wcześniej wymienionych źródeł, pozwala na zmianę stanów epistemicznych, a co za tym idzie, znanych informacji agenta.[27]

Ważną cechą logiki epistemicznej jest jej rozszerzenie o elementy dynamiczne. Dynamiczna logika epistemiczna opisuje zmianę stanów epistemicznych na skutek działań innych agentów *i*. Zachowuje przy tym cechy formalne prezentowanej logiki typu K, czy innych. Znanym i dobrze zbadanym rozszerzeniem jest PAL (Public Announcement Logic) czyli logika publicznego ogłoszenia. PAL formalizuje działanie agenta, który ogłasza wśród innych agentów pewną informację. Informacja ta zmienia stany epistemiczne, zakres wiedzy słuchaczy, może stać się bowiem częścią wiedzy wspólnej (common knowledge).[28]

Rozważmy przykład zbioru symboli *S*:

$$S=\{A_{1,}\, A_{2,}\, A_{3,}\, A_4\}$$

W epistemicznej koncepcji informacji nie ma odniesienia do miary informacji. Natomiast każda informacja ma określone znaczenie, ponieważ dla każdej informacji, utożsamianej z stanem epistemicznym, istnieje dokładnie jeden świat możliwy jako korelat semantyczny. Wynika to z wprowadzonej funkcji *V* w modelu Kripkego.

Chcąc opisać zbiór S w tej teorii, należy uzupełnić posiadane informacje o podmioty, które działają w przykładowej przestrzeni. Przy założeniu, że opis dotyczy tylko jednego podmiotu, należy wskazać zbiór światów możliwych tego agenta, co będzie się wiązało ze

27 por. (Benthem, Martinez 2008)
28 por. (Baltag, Ditmarsch, Moss 2008)

wskazaniem wartościowań dla każdego z tych światów. Częścią jednego, lub więcej światów, będą symbole pochodzące ze zbioru S. Ostatnim krokiem, jaki należy wykonać, jest określenie relacji dostępności dla danego agenta. W ten sposób opis informacji pojedynczego agenta w logice epistemicznej pozwala na wskazanie zakresu wiedzy i dynamiczną jej zmianę.

7 Zakończenie

Przedstawiliśmy pięć ujęć zjawiska informacji. Jak się okazuje, możemy określić dwa dominujące typy opisywania informacji:

- szacunkowy, opierający się na wartości liczbowej informacji
- semantyczny, wynikający z analizy znaczenia informacji.

Każdy z wyznaczonych typów ma swoje zalety. Informacja, dla której można ilościowo określić jej wartość, może być obiektywnie porównana z innymi w danym zbiorze. Ilościowa wartość informacji może być wykorzystana do podjęcia decyzji. Opis matematyczny pociąga także inne konsekwencje, dotyczące na przykład podstaw komunikacji. Na podstawie ilościowej definicji informacji można także próbować opracować formalną teorię informacji.

Natomiast informacja opisywana w kategoriach semantycznych może być wykorzystana na innych polach filozofii. Takie określenie informacji może pozwolić na formalizację w sposób doskonalszy niektórych dziedzin filozofii, przykładowo epistemologii.

Naturalne wydaje się pytanie o możliwość takiej teorii informacji, która łączy zalety każdego z wymienionych typów opisu informacji. Chodziłoby o taką modelową strukturę, która opisuje zarówno ilość informacji i jej znaczenie. Mogłaby być ona wykorzystana przy automatyzacji systemów decyzyjnych, w badaniach nad sztuczną inteligencją czy przy projektowaniu „inteligentnych" komputerów.

Przy odpowiednim poziomie formalizacji, logika informacji, zarówno przy opisie szacunkowym i semantycznym, mogłaby być wykorzystana w projekcie ostatecznego opisu racjonalności ludzkości.

Bibliografia

Baltag A., Ditmarsch H., Moss L. (2008), *Epistemic logic and information update* [w:] *Philosophy of Information*, red. Adriaans P., van Benthem J., Amsterdam, Elsevier, s. 361-457.

Barwise K., Perry J. (1983), *Situation and Attitudes*, Cambridge, MIT Press.

van Benthem J., Martinez M. (2008), *The Stories of Logic and Information*, [w:] *Philosophy of Information*, red. Adriaans P., van Benthem J., Amsterdam, Elsevier, s. 217-280.

Carnap R., Bar-Hillel Y. (1952), *An Outline of a Theory of Semantic Information*, "Technical Report no. 247", Research Laboratory of Electronics, Massachusetts, s. 1-50.

Carnap R. (1966), *Probability and Content Measure*, [w:] *Mind, Matter and Method*, red. P.K. Feyerabend, Minneapolis, s. 248-260.

Carnap R. (1963), *Remarks on Probability*, [w:] "Philosophical Studies", 14(5), s. 65-75**.**

Chaitin G. (2003) , *Two philosophical applications of algorithmic information theory*, [w:] *Proceedings DMCTS'03*, red. C.S. Calude, M.J. Dinneen, V. Vajnovszki, s. 1-10.

Floridi L. (2005a), *Is Semantic Information Meaningful Data?* [w:] "Philosophy and Phenomenological Research", LXX(2), s. 351-360.

Floridi L. (2005b), *Semantic Conceptions of Information* Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/information-semantic.

Grünwald P., Vitanyi P. (2008), *Algorythmic Complexity*, [w:] *Philosophy of Information*, red. Adriaans P., van Benthem J., Amsterdam, Elsevier, s. 289-325.

Israel D., Perry J. (1990), *What is Information?* [w:] *Information, Language and Cognition*, red. Philip Hanson, Vancouver, s. 1-19.

Markowsky G. (1996), *Introduction to Algorithmic Information Theory*, [w:] "Journal of Universal Computer Science", vol. 2 (5), s. 245-269

Meyer J. (2001), *Epistemic Logic*, [w:] *Blackwell Guide to Philosophical Logic*, Massachussets, s. 183-202.

Mortimer H. (1982), *Logika indukcji. Wybrane problemy*, Warszawa, 1982, PWN.

Shannon C. (1948), *A Mathematical Theory of Communication*, "The Bell System Technical Journal", 27 (6-9), s. 379-423, 623-656.